Biochimie 105 (2014) 22–35

Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Research paper

Triplex intermediates in folding of human telomeric quadruplexes probed by microsecond-scale molecular dynamics simulations

Petr Stadlbauer [a], Lukáš Trantírek [b], Thomas E. Cheatham III [c], Jaroslav Koča [b], Jiří Šponer [a, b, *]

[a] Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic
[b] CEITEC – Central European Institute of Technology, Masaryk University, Campus Bohunice, Kamenice 5, 625 00 Brno, Czech Republic
[c] Department of Medicinal Chemistry, College of Pharmacy, University of Utah, Salt Lake City, UT 84124, USA

ARTICLE INFO

Article history:
Received 20 April 2014
Accepted 1 July 2014
Available online 17 July 2014

Keywords:
G-DNA folding
Quadruplex
Triplex
Telomere
Molecular dynamics

ABSTRACT

We have carried out extended set of μs-scale explicit solvent MD simulations of all possible G-triplexes which can participate in folding pathways of the human telomeric quadruplex. Our study accumulates almost 60 μs of simulation data, which is by about three orders of magnitude larger sampling compared to the earlier simulations of human telomeric G-DNA triplexes. Starting structures were obtained from experimental quadruplex structures by deleting either the first or the last strand. The life-times of antiparallel triplexes with lateral and diagonal loops are at least on μs-scale, which should be sufficient to contribute to the folding pathways. However, the triplex states may involve structures with various local deviations from the ideal triplexes, such as strand tilting and various alternative and incomplete triads. The simulations reveal easy rearrangements between lateral and diagonal loop triplex topologies. Propeller loops of antiparallel triplexes may to certain extent interfere with the G-triplexes but these structures are still viable candidates to participate in the folding. In contrast, all-parallel all-anti triplexes are very unstable and are unlikely to contribute to the folding. Although our simulations demonstrate that antiparallel G-triplexes, if folded, would have life-times sufficient to participate in the quadruplex folding, the results do not rule out the possibility that the G-triplexes are out-competed by other structures not included in our study. Among them, numerous possible misfolded structures containing guanine quartets can act as off-path intermediates with longer life-times than the triplexes. Besides analyzing the structural dynamics of a diverse set of G-DNA triplexes, we also provide a brief discussion of the limitations of the simulation methodology, which is necessary for proper understanding of the simulation data.

1. Introduction

Telomeres are specialized, functional DNA-protein structures found at the ends of all eukaryotic linear chromosomes. They help to protect the ends of chromosomes from being treated like damaged DNA needing repair, and they also facilitate complete replication of the chromosome. In vertebrates, telomeres comprise of double-stranded DNA of simple repetitive sequence, d(TTAGGG).d(AATCCC), terminating in a single-stranded G-rich 3′-overhang. Inherent property of the single-stranded G-rich 3′-overhang sequences is their ability to adopt non-canonical DNA structure, namely G-quadruplex. The G-quadruplex formation in telomeric G-rich DNA has been demonstrated both in vitro (reviewed in [1]) and in vivo [2–4]. Recent studies have indicated that telomeric G-quadruplex plays active roles in telomere assembly and integrity and replication/transcription of the telomeric DNA [5–9]. Moreover, as the stabilization of telomeric G-DNA by small molecular weight ligands inhibits activity of telomerase, enzyme responsible for elongation of telomeric DNA and being overexpressed in more than 80% of cancers, the stabilization of the telomeric G-quadruplexes by the ligands has emerged as a potential strategy in the anticancer therapy [10,11].

The most salient structural feature of G-quadruplexes is formation of planar guanine tetrads mediated by Hoogsteen-type guanine–guanine base-pairing. Consecutive tetrads are stacked and stabilized by cations in the central channel. Inherent property

* Corresponding author. Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic. Tel.: +420 541 517 133; fax: +420 541 212 179.
E-mail address: sponer@ncbr.muni.cz (J. Šponer).

http://dx.doi.org/10.1016/j.biochi.2014.07.009

of the telomeric G-quadruplexes is their structural polymorphism. Biologically relevant monomolecular quadruplexes can adopt various topologies depending on their sequence and surrounding conditions [12–18]. G-strands can be either parallel or antiparallel, and are connected via loops, classified into lateral, diagonal or double-chain reversal (propeller) types, which gives rise to parallel, hybrid and antiparallel topologies [10,15,19–28]. Moreover, additional diversity is brought by *syn* or *anti* orientation of the glycosidic torsion angle χ of guanosines in the G-stem. The possible *syn*/*anti* patterns are constrained by strand orientations and affected also by presence or absence of flanking nucleotides at the beginning/end of a sequence [29,30]. Thus far, six distinct folding topologies of telomeric G-quadruplex were reported [12–16,27].

While the diverse telomeric G-DNA structures have been rather well characterized experimentally, much less is known about folding kinetics and folding pathways for the individual G-quadruplexes. Importance of the kinetic information has emerged recently along with the experimental data suggesting that number of physiologically relevant processes involving telomeric G-DNA folding/unfolding such as chromatin remodeling, replication, transcription, protein binding or sequestration of non-coding RNA might be under kinetic rather than thermodynamic control [31,32]. However, the experimentally observed folding rates and the number of folding phases may be affected by the experimental setups involving the cation and buffer conditions, the methods used to interrogate the folding, or whether the G-DNA construct used for the investigations has flanking nucleotides [32–36]. Yet, regardless of the discrepancies, these studies have collectively indicated that telomeric G-DNA folds on millisecond to second time scale via multiphasic processes involving intermediates [32–34,37–39]. Recently, the pathway for unimolecular quadruplex folding of human telomeric sequence was simultaneously monitored by several kinetics experiments, providing the so far most comprehensive insight into the G-DNA folding process [38]. This study suggested that the folding process consisted of phases with times scales ranging from ms to 1000 s.

The triplex structure was suggested as one of the most plausible intermediates in folding pathway of human telomeric quadruplex [37,40–45]. Yet, due to the limited temporal and structural resolution of the experimental techniques used to monitor G-DNA folding as well as due to the model-dependent nature of kinetic data interpretation, the structure of the intermediate (triplex) has remained speculative and the data have not provided any deeper insight into molecular events taking place along the folding pathway. Actually, while some studies suggested triplexes acting as on-pathway intermediates, other studies suggest triplexes acting as long-living off-pathway intermediates [46].

In principle, although limited by force field accuracy and affordable time scale, computer simulations (molecular dynamics, MD) can complement experimental data and provide insights into selected aspects of G-DNA folding [47–54]. For example, standard simulations suggested participation of straightforward vertical strand-slippage movements in late stages of formation of parallel tetrameric all-*anti* G-DNA stems [47,48]. In contrast, antiparallel and hybrid G-stems are unable to undergo direct vertical slippage of their strands due to steric conflicts between *syn* and *anti* oriented bases [48]. Thus, their rearrangements rather occur via strand binding-unbinding processes coupled with *syn* − *anti* dynamics of the individual nucleotides. This observation would indirectly support presence of the triplex intermediates.

Standard (unbiased) simulations allow observing only those motions that a real molecule would sample on the simulation time scale, which is presently in the microsecond range. Standard simulations, therefore, are too short to visualize full folding of G-quadruplexes. This limitation can be partially alleviated by investigation of a diverse spectrum of starting structures that could occur along the folding pathway [47,48]. The alternative option is utilization of enhanced sampling simulations which use various approaches to overcome the energy barriers. Steered dynamics, temperature replica exchange molecular dynamics (T-REMD) and metadynamics have been used to investigate G-DNA unfolding [50–53]. However, enhanced sampling simulations are not fully equivalent to standard simulations and their outcome may be affected by the specific method of enhanced sampling [55]. T-REMD accelerates crossing over enthalpic barriers but cannot increase sampling of entropy-driven processes and therefore changes the entropy/enthalpy balance. Metadynamics assumes that the unfolding process can be sufficiently completely described by only a few collective variables providing a coarse-grained description of all the important slow modes. If the real process includes additional slow motions that are orthogonal to the defined collective variables (which may happen for complex systems), the metadynamics description becomes inaccurate and incomplete. Metadynamics can be unsatisfactory when different stages of the folding path or different routes of a multi-pathway process are dominated by different types of slow motions, requiring different collective variables. The so far most advanced enhanced sampling G-DNA study reported unfolding of the human telomeric hybrid-1 quadruplex using four-dimensional bias-exchange metadynamics using four collective variables [54]. The study also illustrates (as discussed by its authors) limitations of the metadynamics, such as the lack of acceleration of *syn*/*anti* guanine reorientations not included in the collective variable set. This precludes sampling of alternative folds and of misfolded structures. Another study combined metadynamics simulations with NMR experiments which provided evidence for formation of the triplex structure by truncated sequence of the thrombin binding aptamer (15-TBA) DNA [50]. Note that none of the currently available studies could simulate true folding of G-DNA. All studies have been initiated from folded structures and thus capture unfolding. The accelerated disruption of a folded molecule does not necessarily capture the folding pathway and, in addition, the results may be influenced by the specific enhanced sampling method. For example, the T-REMD [53] and metadynamics [50] models of 15-TBA G-DNA unfolding differ visibly. As demonstrated for very simple RNA tetraloops, achieving folding from the unfolded state would be considerably more difficult [56]. Therefore, despite some successes, we remain far from an exhaustive computer description of true G-DNA folding processes.

An important suggestion emerging from recent standard simulations is the possibility that folding pathways of G-DNA molecules include at the atomistic level numerous intermediates which cannot be resolved by the existing experimental techniques, resulting in extremely multi-pathway process [48]. Even a capture of a single ion by a G-DNA stem likely proceeds via a number of distinct micro-routes [57]. High complexity has been explicitly revealed by standard simulations in folding studies of the smallest DNA hairpins that fold on microsecond time scale [58]. Thus, the human telomeric G-quadruplex folding pathways suggested by the recent experiments could potentially be oversimplified by the resolution of available methodologies. G-quadruplex folding may resemble folding funnels suggested for proteins and RNA [59–62].

Polymorphism of folded G-quadruplex structures in the thermodynamics equilibrium is well established [12–18,21,25,27]. However, diversity of the structures occurring during folding processes may be much larger. A single G-DNA folding strand can in principle adopt 2^N different combinations of *syn* and *anti* G orientations (N is the number of guanines involved in the G-DNA

tetrads), i.e., there are 4096 unique *syn*/*anti* guanine combinations for a DNA strand of the human telomeric quadruplex. Different initial *anti*/*syn* substates may result in different atomistic folding routes with diverse off- and on-pathway intermediates. It is actually difficult to imagine that any presently available computational method could provide converged sampling of the G-DNA unfolded state, i.e., the equilibrium distribution of the diverse *anti*/*syn* substates. Many of these *anti*/*syn* combinations can then form various stable non-native individual structures throughout the folding process before reaching the final equilibrium [48]. Any G-quadruplex forming strand can in principle adopt altogether 26 distinct G-quadruplex topologies, differing in strand directionality and type of the loops [22,30]. Although not all of them are energetically feasible, at least five three-quartet (and in addition one two-quartet) topologies have already been observed for the human telomeric quadruplex (see above) as dominant equilibrium structures in various atomistic experiments [12–16,27,29,35,36,49,63]. Further, the G-rich sequence can potentially form quadruplexes with incomplete number of tetrads, including structures with a slipped strand (actually, the two-quartet quadruplex PDB: 2KF8 [12] falls into this category). For any quadruplex fold with three or more tetrads, eight different alternative structures with loss of one tetrad can be obtained by slipping one of the strands in either 3′ or 5′ direction and adjusting the *syn*/*anti* pattern [48]. Therefore, the diversity of structures which a given G-rich DNA strand can potentially adopt is enormous. While only a small fraction of these possible structures have been seen in atomistic experiments, many more can be temporarily populated during the G-quadruplex folding processes as on-pathway and off-pathway intermediates.

In the present study, we provide insights into properties of potential triplex intermediates, which can participate in folding of human telomeric quadruplex. We simulate six different variants of monomolecular triplexes that are plausible candidates for folding intermediates of parallel, antiparallel and hybrid (3 + 1) topologies of human telomeric G-DNA. As our simulations are limited by affordable time scale, reaching microseconds in recent studies [48,64–66], we are not able to follow the whole folding process starting from a single-stranded randomly structured DNA. Instead, we start from structures based on available atomistic experiments and investigate their structural dynamics. Our goal is to obtain insight into (in)stability of various triplex topologies and estimate their lifetimes. Our results indicate which human-telomeric quadruplexes could be formed via triplex intermediates and where it is less likely. The time-scale of our study (accumulating in total 56.5 μs) is by about three orders of magnitude larger compared to the earlier simulations of triplexes relevant for folding of human telomeric G-DNA [43,44] and is sufficient to see their rearrangements. Limitations of our study are in more details discussed below.

2. Material and methods

2.1. Preparation of three-layer triplex models

To build up the triplexes, we have used four monomolecular three-tetrad quadruplex structures: the X-ray structure of parallel-stranded quadruplex d[$A(G_3T_2A)_3G_3$] (PDB: 1KF1) [14], the NMR solution structure of a 3 + 1 (hybrid) quadruplex d[$T_2(G_3T_2A)_3G_3A$] (PDB: 2GKU, the first frame) [25], the NMR solution structure of a 3 + 1 (hybrid) quadruplex d[$T_2A(G_3T_2A)_3G_3T_2$] (PDB: 2JPZ, the first frame) [16], and the NMR solution structure of antiparallel (basket type) d[$A(G_3T_2A)_3G_3$] quadruplex (PDB: 143D, the first frame) [13].

The quadruplex structures were truncated to a triplex by deleting either the first or the last G-strand and the adjacent loop, resulting is eight structures representing six different triplex topologies (Fig. 1). Whenever possible, one flanking residue at each end of the triplex was preserved. Thereby, two of the models were parallel stranded with two propeller loops (based on 1KF1, models **A** and **B**, sequences d[$A(G_3T_2A)_2G_3$] and d[$A(G_3T_2A)_2G_3T$]), one antiparallel with a diagonal loop followed by a lateral loop (based on 143D, model **C**, d[$A(G_3T_2A)_2G_3$]), one with swapped order of the loops (based on 143D, model **D**, d[$A(G_3T_2A)_2G_3T$]), two were antiparallel with two lateral loops (based on 2GKU, model **E**, d[$A(G_3T_2A)_2G_3A$] and on 2JPZ, model **F**, d[$A(G_3T_2A)_2G_3T$]), one antiparallel with a propeller loop followed by a lateral loop (based on 2GKU, model **G**, d[$T(G_3T_2A)_2G_3T$]), and one with swapped order of the loops (based on 2JPZ, model **H**, d[$A(G_3T_2A)_2G_3T$]). Note that the structures **A** and **B**, and **E** and **F** have the same topology but differ in some structural details. We did not include the recently discovered antiparallel quadruplex with middle propeller loop (PDB: 2JMB) [27], as the resulting triplexes would have topologies identical to our models **G** and **H**. The three-triad triplex relevant for the two-quartet 2KF8 structure (strands 2–4) would be topologically equivalent to the structure **C**, though the upper triad has one *syn* to *anti* flip due to other specific interactions. Our study thus includes all possible three-triad triplexes relevant for a folding of all currently known human telomeric G-DNA folds.

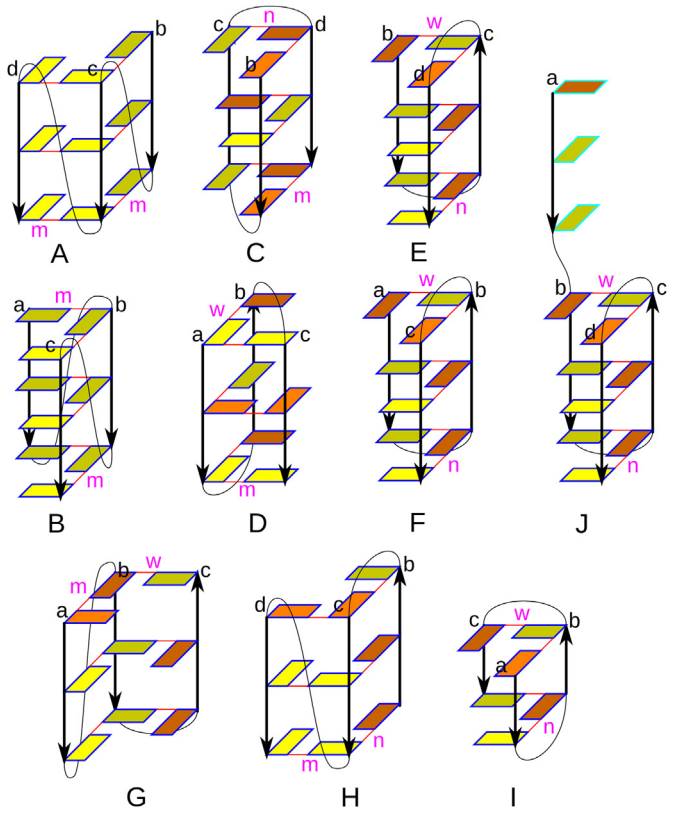

Fig. 1. Structural schemes of starting structures used in our simulations. Triplexes are shown as follows: deoxyguanosine residues are either yellow (*anti* orientation) or orange (*syn* orientation) rectangles; darker residues are in the back. Backbone is black and 5′ → 3′ direction is denoted by the arrows. Loops are depicted by thin black curves. *c*WH pairing is labeled by solid red lines. Ions are not shown. Coloring of the edges of the rectangles in the structures marks residues with approximately the same normal vector of their respective base plane. The structures are marked by letters **A**–**J** which are also used throughout the whole paper. Letters **a**–**d** denote G-strands (G-stretches) and mauve letters indicate type of groove (m – medium, n – narrow, w – wide). Supporting Material provides PDB files of all starting structures except for the molecule **I** and also numerous interesting structures occurring in the course of the simulations.

2.2. Triad and strand nomenclature

Further in the text, the first (upper) triad refers to the triad that contains deoxyguanosine closest to the 5′-end (that is G2 in all triplexes), the second triad is the middle one and the third the bottom one, according to the starting triplex structure. Thus the first triad in a particular triplex is not necessarily equivalent to the first tetrad in the corresponding quadruplex molecule. However, G-strands are labeled consecutively **a**, **b**, **c**, and **d** according to the original quadruplexes.

2.3. Initial positioning of the bound ions

We have tested three different initial layouts of cations in the triplex channel for all but parallel systems: (i) one cation placed between the first and the second triad; (ii) one cation placed between the second and the third triad; (iii) two cations, meaning full initial occupation of the channel. The parallel triplexes have only been simulated with two channel cations.

2.4. Two-triad triplex

We have studied the recently suggested d[$G_2T_2G_2TGTG_2$] triplex [50], derived from the thrombin binding aptamer (15-TBA) quadruplex by removing its last G-strand. It forms a two-triad structure with two lateral loops and one cation between the triads (Fig. 1, structure **I**, we have taken the structure provided in Ref. [50] as the starting structure). We have also simulated two-triad triplex models that combined loop geometry either as provided in Ref. [50] or taken from the 15-TBA NMR structure (PDB: 148D, the eighth frame, the first and the second loop) [67] with geometry of the first two triads obtained from our simulations of the molecules **E** or **J**. (for details see Supporting material) The model structures were built up using Swiss-Pdb Viewer [68]. The starting structures included one channel cation.

2.5. Modeled triplex with single strand

We have also simulated an antiparallel triplex with two lateral loops with attached unfolded fourth G-strand (Fig. 1, structure **J**). The model was prepared from the structure of 2GKU by manual adjustment of some backbone dihedral angles of loop residues in order to flip the first G-strand above the remaining triplex, abrogating the structure of the propeller loop.

2.6. Water and ionic conditions

The solute molecules were placed in a truncated octahedral box of water molecules with minimal distance between the box border and solute surface of 10 Å at all points. TIP3P water model [69] was used in all NaCl simulations, SPC/E model [70,71] was applied in KCl simulations (Table 1, Supporting Table SI). All simulations of three-triad triplexes were carried out in the presence of 0.15 M excess salt. Simulations of the two-triad triplex were performed either in 0.07 or 0.15 M excess salt (Supporting Table SI). Standard ion parameters were used [72].

It is well established that the final equilibrium of folded G-DNA molecules depends on the type of the ions [13,14,20,25,73]. However, microsecond-scale simulations are too short to be radically sensitive to differences between Na^+ and K^+ ions. For the purpose of the present simulations, Na^+ and K^+ ions are assumed to provide essentially equivalent results. The difference between effect of the ion type in MD simulations and in equilibrium TD experiments of G-DNA and the reason why the cation type does not have any profound effect on the behavior of the μs-scale simulations are in detail discussed in our recent study [48] on p. 7137. In fact, experiments monitoring the transition of the basket structure to the hybrid structure upon the $Na^+ \rightarrow K^+$ exchange reveal that while the ion replacement in the quadruplex is completed on a time scale of ~250 μs, the basket structure remains stable even after dozens of second [74]. Thus, all known human telomeric quadruplex folds should be (as single molecules) compatible with both monovalent ions. The ion type, other environmental effects and the flanking sequences modulate the relative free energies of the different folds and that is why a given equilibrium experiment shows measurable populations of only some of the structures. However, all the folds are supported by both cations *per se* (i.e., they would be present if the competing structure is eliminated) and if formed in a single molecule event they would have long lifetimes.

Solvation and addition of ions were done using xLEaP module of AMBER 12 [75].

2.7. MD simulations

MD simulations were carried out with the parmbsc0 [76] version of the Cornell et al. force field [77], which has been the default AMBER force field for DNA since 2007. Some simulations were repeated by adding also the 2012 parmχ_{OL4} modification to bsc0. Parmχ_{OL4} improves the shape of the *syn* region of the χ torsional potential of the force field and affects the balance and kinetics between the *syn* and *anti* χ regions [78]. We tested also addition of the 2013 parm$\varepsilon\zeta_{OL1}$ modification, which refines the ε/ζ torsion potentials, improving indirectly also balance of the noncanonical α/γ backbone substates in G-DNA [79] (Table 1, Supporting Table SI). The later modifications are not essential to achieve a basic stability of DNA simulations and are still under testing. While this paper has been under review, the bsc0 + χ_{OL4} + $\varepsilon\zeta_{OL1}$ combination of the force field refinements has been suggested by the AMBER code developers as the optimal combination to describe DNA systems. Nevertheless, since in our study we do not analyze details of the conformations, we assume that the bsc0 is sufficient to get the correct basic results.

The simulations were performed with the CUDA version of pmemd module of AMBER 12 [75,80–82]. Periodic boundary conditions were used, and electrostatic interactions were calculated by the particle mesh Ewald method [83,84] with the non-bonded cutoff set to 9 Å. The SHAKE algorithm [85] was applied to bonds involving hydrogens, and a 2 fs integration step was used. Pressure was held constant at 1 atm and temperature at 300 K, using the Berendsen weak-coupling thermostat [86]. Equilibration protocol is described in Supporting material (Section 9.1). Trajectories were processed with the ptraj module of AMBER and visualized in VMD [87].

3. Results

3.1. G-DNA-like vs. alternative triads

In the following paragraphs, we describe multiple simulations of different potential triplex intermediates that could participate in G-DNA folding. Besides conventional Hoogsteen triads (*cis*-Watson-Crick – Hoogsteen, *c*WH pairing), we have often observed competing symmetrical triads lacking the bound ions and stabilized by bifurcated *c*WH pairing. We call them *triangle triads* and we suggest that they may participate in the folding processes, as noticed also by Mashimo et al. [43]. Other intermediate triads frequently encountered in our simulations during strand transitions include those with one stable *c*WH base pair and the third base either placed along the edge of the pair or forming *t*WW

Table 1
List of simulations of three-triad triplexes.

Simulated structure[a]	Loops[b]	Channel cations[c]	Simulation length (ns)	Simulated structure[a]	Loops[b]	Channel cations[c]	Simulation length (ns)
A	p → p	2	100	F	l → l	1	1000
A[d]	p → p	2	250	F	l → l	1	1000
B	p → p	2	200	F	l → l	2	1000
B[d]	p → p	2	250	F[d]	l → l	1	1000
C	d → l	1	1000	F[d]	l → l	1	1000
C	d → l	1	1000	F[d]	l → l	2	1000
C	d → l	2	1000	G	p → l	1	1000
C[d]	d → l	1	1000	G	p → l	1	1000
C[d]	d → l	1	1000	G	p → l	2	1000
C[d]	d → l	2	1000	G[d]	p → l	1	1000
D	l → d	1	1000	G[d]	p → l	1	1000
D	l → d	1	1000	G[d]	p → l	2	1000
D	l → d	2	1500	G[e]	p → l	1	1000
E	l → l	1	1000	G[e]	p → l	1	1000
E	l → l	1	1000	G[e]	p → l	2	1000
E	l → l	2	5000	H	l → p	1	1000
E[d]	l → l	1	1000	H	l → p	1	1000
E[d]	l → l	1	1000	H	l → p	2	1000
E[d]	l → l	2	1000	J[e]	0 → l → l	2	5000

[a] See Fig. 1; structures were simulated under parmbsc0, TIP3P and NaCl excess salt conditions, unless stated otherwise.
[b] Order of loops in the initial triplex; l – lateral, p – propeller, d – diagonal, 0 – perturbed loop.
[c] Number of cations placed in the channel at the beginning of simulation.
[d] bsc0$\chi_{OL4}\varepsilon\zeta_{OL1}$ force field.
[e] bsc0χ_{OL4} force field.

(*trans*-Watson-Crick – Watson-Crick) base pair (Fig. 2). The latter has been described by Stefl et al. [47].

3.2. Parallel stranded all-anti triplexes are unstable

Four simulations of parallel all-*anti* triplexes (models **A** and **B**) showed swift loss (within ten ns) of the triplex structures (Fig. 3). More detailed description is given in the Supporting material (Section 10.1). The present results are in agreement with the decade-old study by Stefl et al. showing (albeit with an older force field version and not considering the loops) instability of four-layer parallel stranded G-triplexes [47]. Thus, well-structured parallel-stranded all-*anti* triplexes are unlikely to participate in folding/formation processes of G-DNA. It is more likely that the folding pathway of all-parallel all-*anti* structures includes various four-stranded intermediates with incomplete number of tetrads progressing towards the final structure via strand slippage processes [48].

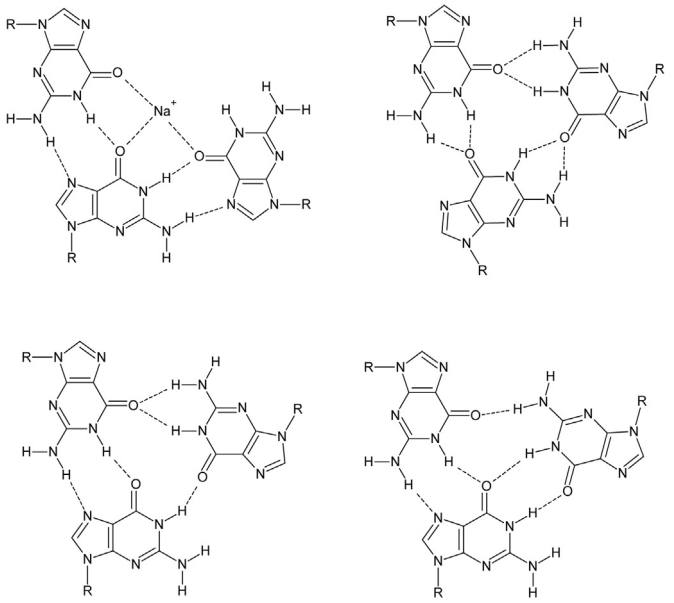

Fig. 2. Observed triad variants: ideal Hoogsteen triad (top left), ideal triangle triad (top right), triad with one *c*WH base pair and the third base bound to the edge of the base pair (bottom left) and triad comprising *c*WH and *t*WW base pairs (bottom right). R depicts the sugar-phosphate moiety.

3.3. Antiparallel triplexes with a diagonal loop are stable albeit imperfectly paired

Nine simulations of antiparallel triplexes with diagonal and lateral loops (structures **C** and **D**) show much larger propensity to keep triplex-like structures for a considerable amount of time compared to the parallel all-*anti* triplexes. The simulations, however, indicate that such triplexes would be imperfect. Although the simulations preserved the overall triplex arrangement, the structures were locally perturbed, in a different way in different simulations. These deformations, however, should not prevent capture of the fourth G-tract. Thus we suggest that such triplexes, once formed, should have sufficient life-times to allow further progression of the folding. The triplex-like structures are predicted to extensively interact with ions.

Let us now describe the simulations in more detail. When assessing simulations, we need to consider that the simulations are stochastic. Therefore, differences between the individual simulations may often be due to their genuine stochasticity and not due to the different simulated starting structures. Thus, for interpretation of our results, we primarily consider common qualitative features seen in multiple simulations.

The triplex **C** with diagonal → lateral order of the loops showed a common feature in all three parmbsc0 simulations, namely, tilt of its first strand **b** with respect to the duplex of strands **c** + **d**. The third triad was always disrupted due to the strand **b** tilt. Eventual additional changes are described below. Importantly, as noted above, after the initial rearrangements the structures remained stable till the end of the 1 μs simulations.

The structure initially with a cation between the second and the third triad captured another cation during equilibration, however,

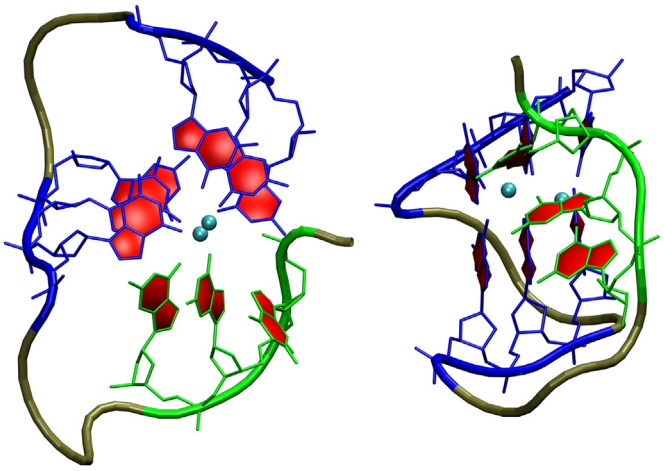

Fig. 3. Cross-like structure formed in simulations of all-*anti* all parallel triplex **A** after 2 ns. Strand **b** (green) is roughly perpendicular to duplex **c** + **d** (blue). Channel cations are cyan, loop backbone is tan, only deoxyguanosine residues are shown. Hydrogen atoms are not shown. Top view on the left, side view on the right. The triplex structure is quickly lost.

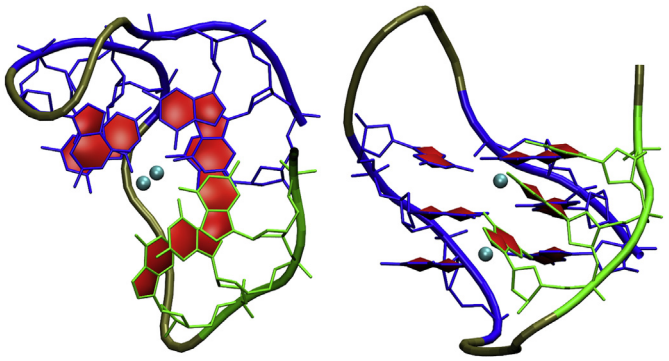

Fig. 4. Tilt of the strand **b** (green) from the duplex of **c** + **d** (blue) in the diagonal-lateral loop triplex **C** as seen in the simulation starting with a cation between the second and the third triad. Channel cations are cyan, loop backbone is tan. Top view – left, side view – right. Only deoxyguanosine residues are shown. Hydrogen atoms are not shown. This triplex-like arrangement is stable.

the strand **b** tilted. The third triad opened between 90 and 130 ns, and ultimately broke up at 450 ns. The two cations and the first two triads (slightly affected by the initial tilt) remained intact (Fig. 4, Supporting Fig. S1). The simulation with the cation initially bound between the first and the second triad revealed the same overall picture, although speeded up by absence of a cation stabilizing the third triad. This triad adopted triangle shape (cf. Fig. 2) during equilibration and was permanently lost since 50 ns of the simulation. The first and the second triad were maintained and the ion binding site between them was almost continuously occupied, with six ion exchanges with the bulk. Finally, the strand **b** of the structure with two initially bound cations tilted during the first 2 ns. Further, the first deoxyguanosine of this strand (G2, the first triad) flipped into *anti* orientation at 15 ns, which prevented its correct pairing with the duplex of strands **c** + **d**. Both cations remained bound in the visibly perturbed channel for the next 300 ns. Then one ion was lost and during the rest of the simulation the channel was occupied just by one cation with eleven exchanges with the bulk. Since the tilt of the strand **b** was larger than in the preceding two simulations, essentially all triads were broken. The tilted structure, however, remained stable till the end of the simulation.

Three simulations with the bsc0$\chi_{OL4}\varepsilon\zeta_{OL1}$ force field revealed very similar behavior as bsc0 data with propensity of the strand **b** to tilt (Supporting material – Section 10.2).

Simulations of the molecule **D** with lateral → diagonal order of the loops can be described as follows. The system with starting cation between the first two triads had perturbed third triad until 340 ns, when a second cation entered the channel and the third triad was completed. At 530 ns, the first G of the strand **a** adopted a *t*WW interaction with G in the strand **c** and weakened its interaction with G in the strand **b** (Supporting Fig. S2). Interestingly, the slightly perturbed first triad with only two guanine oxygens aiming toward the central channel was able to maintain the channel ion binding site between the first and the second triad. The slightly perturbed triplex with two bound cations was then entirely stable till the end (1000 ns). The triplex possessing initially one cation between the second and the third triad caught another cation during the equilibration. At 35 ns the strand **c** became tilted. After the next 100 ns, the initially bound cation left the channel while tilting of the strand **c** increased. The deformed triplex did not undergo any further changes. The third simulation of molecule **D** with two initial cations showed the largest deviations and is described in the Supporting material (Section 10.2).

3.4. Antiparallel triplex with two lateral loops is stable, but reveals competition between Hoogsteen and triangle triads and transitions from lateral to diagonal loops

Interesting rearrangements occurred in the simulation of the model **E** with one cation initially bound between the second and the third triad. The strand **b** changed its binding partner from the strand **c** to the strand **d**, transforming the former lateral loop to a diagonal loop (see the Supporting material – Section 10.3 – for detailed description) (Fig. 5). This also means replacement of the wide groove by the medium groove. Such horizontal strand binding exchange preserves compatibility of the *syn*/*anti* distribution of the antiparallel triplex (in contrast to a vertical slippage of the strands). Such strand relocations could commonly occur during dynamics of real triplexes and switches between lateral and diagonal loops during dynamics and folding of single G-DNA molecules appear straightforward.

The sequence of events in the simulation starting with two cations in the channel resembled the preceding simulation and we prolonged this simulation to 5 μs (Supporting Fig. S3). The original triplex was stable until 600 ns, when the bottom cation left the channel. After the next 60 ns the second cation was expelled and a triplex with two triangle triads and one imperfect triangle triad (cf. Fig. 2) was formed within the next 20 ns. At 700 ns, an unsuccessful attempt of a cation to enter the triplex occurred. Lifetime of the triangle triplex was about 500 ns, until its fluctuations facilitated capture of a cation. Then the strand **b** ultimately changed its binding partner from the strand **c** to the strand **d**. The cation resided between the first and the second triad making them perfectly Hoogsteen paired and the loop between the strands **b** and **c** became diagonal. The third triad remained in the triangle arrangement (Fig. 5). The cation left the triplex at 1670 ns and the triplex again adopted the triangle triads structure. Then, till the end (5 μs) of the simulation, we observed competition between the ion-stabilized Hoogsteen triads and triangle triads, with successful and unsuccessful ion binding events. This simulation thus confirmed easy transitions between lateral and diagonal loops in the triplexes, although we did not see return of the lateral loop, i.e., the fluctuations occurred between the diagonal loop and intermediate arrangements. The Hoogsteen and triangle triads may dynamically compete while preserving the overall triplex topology for very long time.

Note that during the formation of the triangle-triad triplex the difference between the lateral and diagonal loops essentially vanishes, since a loop spanning across a triangle triad is neither lateral nor diagonal; its topology is intermediate. Nevertheless, if the middle (or eventually the distant) triad is Hoogsteen, we can formally classify the triangle-triad loop according to the pairing of this Hoogsteen triad.

The last simulation where the channel cation was initially placed only between the first and the second triad showed six ion-release-and-capture events in the initially occupied ion binding site. The Hoogsteen pairing in triads was well maintained, competing with ~10% population of triangle triads. After 600 ns, during one ion exchange event with no ions in the channel, the triplex unfolded to a duplex of strands **c** + **d** and the diverting strand **b.** There was a subsequent unfolding-refolding dynamics period which lasted ~90 ns. The strand **b** was at one moment completely detached, having no hydrogen bonds or stacking with the **c** + **d** duplex. The **c** + **d** duplex remained intact. Then the triplex reformed and captured a cation. However, the strand **b** remained partially tilted with respect to the duplex. At 980 ns the first deoxyguanosine of the strand **b** flipped to *anti* (Supporting Fig. S4). Still, this structure can be considered as a perturbed triplex-like arrangement.

Simulations of the molecule **F** were consistent with the above results. They confirmed propensity to change the lateral loop into the diagonal loop and the overall stability of the triplexes. Similar results were obtained when parmbsc0$\chi_{OL4}\varepsilon\zeta_{OL1}$ force field parameters were applied to the molecules **E** and **F** (Supporting material – Section 10.3).

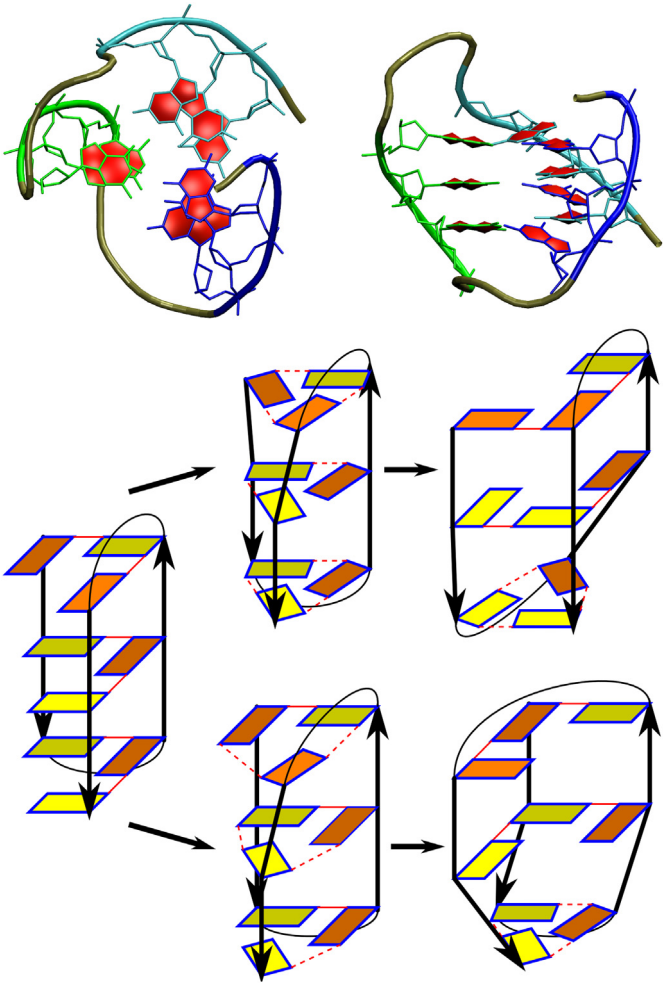

Fig. 5. Top: Top view and side view of the triangle-triad triplex observed in the simulation of the molecule **E**. G-strands **b**, **c**, **d** are colored consecutively blue, green and cyan, loop backbone is tan. Only deoxyguanosine residues are shown. Hydrogen atoms are not shown. Bottom: Two scenarios of strand relocation observed in our simulations. The upper goes through the triangle triplex (shown in the top) to two *c*WH triads and a triangle triad adjacent to the newly formed diagonal loop (two simulations of the molecule **E**). The lower pathway goes through an imperfect triangle triplex, capable of holding channel cations, to an imperfect triangle triad and two *c*WH triads adjacent to the newly formed diagonal loop (one simulation of the molecule **F**). The schemes are visualized as in Fig. 1. In addition, non-G-DNA-like hydrogen bonding is depicted by dashed red lines.

3.5. Propeller loops may destabilize triplexes

Three simulations of the triplex **G** indicate that interactions between the G-strands and the propeller loop as well as internal instability of the propeller loop may to certain extent interfere with the formation and stability of such triplex intermediates.

The first simulation with one initially bound cation in the channel resulted into stable triplex structure with complete second and third triad. The GG base pair from the broken first triad (disrupted due to some tilting of the strand **a**) was stacked on the adjacent triad and stabilized by a fully occupied ion binding site between the triads and two weaker ion binding sites. More details are given in the Supporting material (Section 10.4).

The other two simulations, however, lost the triplex structure. At 4 ns of the second simulation starting with one bound cation, the third triad split into a base pair and a free nucleotide. The free nucleotide bulged out into the solvent and flipped from *syn* to *anti* orientation. The perturbed triplex immediately captured a cation so that both binding sites were occupied. This imperfect structure existed for the next 630 ns. Then the bulged-out deoxyguanosine with altered glycosidic orientation returned to the triplex and formed a misfolded triad, causing the triplex to gradually degrade. At 900 ns, the triplex lost its last channel cation and was split into a duplex of the strands **b** + **c** and the strand **a**, which was H-bonded with the propeller loop residues. The locally misfolded structure remained in this arrangement till the end of the simulation (1000 ns). This simulation indicates that interactions between the G-strands and the propeller loops may obstruct (or compete with) formation of the native interactions during the G-DNA folding process.

The triplex simulated with initially two cations in the channel showed the largest changes. It was stable until 75 ns when the first deoxyguanosine of the strand **a** turned *anti*, so it became incompatible with the native quadruplex pairing. This led to disruption of the first triad and loss of the adjacent cation. Then the strand **c** detached. Although it returned back to the duplex in the next 8 ns, it remained tilted. At 425 ns the last cation left the structure and the structure split into a duplex of strands **b** + **c** and the separated strand **a**. The first two guanines of the strand **a** were sandwiched by propeller loop residues. The third deoxyguanosine was fluctuating nearly freely in the bulk solvent while being flipped to *syn*. During the next 130 ns the detached strand turned upside down so that the propeller loop became lateral-like. At 700 ns, the first deoxyguanosine of the strand **a** turned back to the *syn* region. However, the strand **a**, due to its above-noted reorientation, could not bind to the duplex, because of its incompatible *syn-anti* pattern. A cavity, occupied by a cation, was formed inside the molecule. No further changes occurred during the next 440 ns (Fig. 6, Supporting Fig. S5). Such structure, if formed, would represent one of the misfolded off-pathway intermediates (see Introduction – Section 1).

Simulations of the molecule **H** showed propeller loop instability in one of the three simulations. Details of all three simulations are given in the Supporting material (Section 10.4).

Despite that the above-described simulations indicate that formation of propeller loops during G-DNA folding may be more difficult (less probable) than formation of lateral/diagonal loops, considering the sampling limitations and force field approximations we still suggest that such triplexes can occur during the folding process. They are definitely considerably more viable than the absolutely unstable all-*anti* parallel stranded triplexes – see Section 3.2. Since behavior of triplexes with propeller loops was the most complex and thus potentially sensitive to the details of the force field, we repeated all three simulations of the system **G** with bsc0χ_{OL4} and bsc0$\chi_{OL4}\varepsilon\zeta_{OL1}$ force field modifications. In these simulations, the propeller loops were stable (further details are in Supporting material – Section 10.4). The simulations, among other things, confirmed competition between Hoogsteen and triangle triads and capability of the lateral loop to change to the diagonal arrangement, as in simulations of the system **E**. Unfortunately, with μs scale sampling limitation and considering the approximate nature of the force fields, we are unable to further refine our description of the propeller-loop triplexes. Analysis of antiparallel triplexes with propeller loops thus remains less certain compared to the other triplexes analyzed above.

3.6. Triplex from thrombin binding aptamer shows μs-scale lifetime

We have carried out fourteen independent simulations of the two-triad triplex based on the truncated 15-TBA aptamer (molecule **I**; Supporting Table SI). Four structures were lost (i.e., the triplex irreversibly disintegrated) within the first 100 ns, four triplexes were lost within 100–200 ns and two additional ones in the remaining part of the simulation. In four simulations the triplex was maintained till the end (1 μs). Note that out of the four long-lived structures, three were simulated using the latest parmbsc0-$\chi_{OL4}\varepsilon\zeta_{OL1}$ force field version. Only three triplex structures survived 50 ns and one survived over 150 ns with parmbsc0 force field. However, we presently do not have enough data to decide which force field variant is more realistic for this specific system and thus we consider all fourteen simulations as equally significant. Our simulation results indicate that the 15-TBA triplex, once formed, could have sufficient lifetime to promote further folding. Nevertheless, when considering the simulations alone, we would not predict the (truncated) triplex to be the dominantly populated structure, since the triplexes are progressively lost upon prolongation of the simulations. This indicates that the relation between theory and experiment is more complicated than suggested in Ref. [50] and we have thus performed several additional investigations.

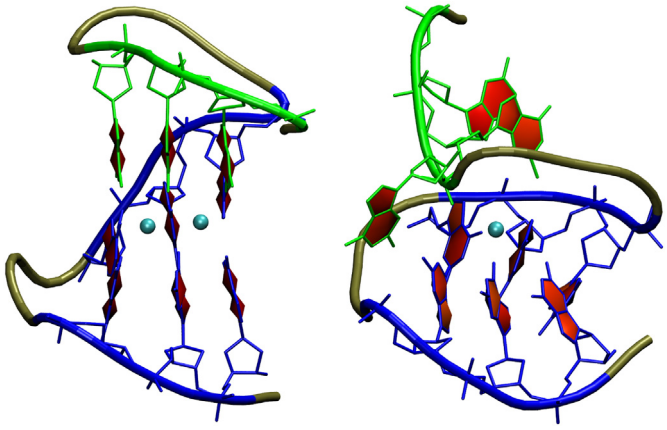

Fig. 6. Change of orientation of the strand **a** (green) with respect to the strands **b** and **c** (blue) in the propeller-lateral loop molecule **G** (simulation with initially two internal ions). The initial propeller loop between **a** and **b** (left, 0 ns) is transformed to lateral-like loop (right, 1000 ns) during the simulation. Loop backbone is tan, channel cations are cyan, only deoxyguanosine residues are shown. Hydrogen atoms are not shown.

First, we monitored consistency of the primary NMR data from Ref. [50] with the simulations (see the Supporting material – Section 10.5 – for further details). When the triplex is stable in simulations, it is fully consistent with the primary NMR data [50]. When the triplex structure is lost, the simulation becomes entirely inconsistent with the NMR data. Ref. [50] concluded that the triplex is stable in simulations, which contrasts our results. This difference can, however, be easily explained since stability of the triplex simulations in Ref. [50] has been judged (besides the metadynamics quickly unfolding the structure) using two short (50 and 100 ns) standard simulations. Thus, instability of the triplex in long unbiased simulations was not revealed. In contrast, we performed fourteen 1 μs simulations, thereby increasing the unbiased sampling by almost two orders of magnitude. We emphasize that we in no case question the primary NMR data supporting the triplex structure [50]. However, we suggest that long simulations do not support the triplex as unambiguously as reported in Ref. [50]. We nevertheless suggest that the MD simulation and experimental NMR results can be brought into accord by assuming that the simulation methodology (i.e., the simulation force field) somewhat underestimates the absolute free energy stability of the triplex, its lifetime and its population (see discussion of the limitations of the MD technique – Sections 4.2 and 4.3). For further discussion of limitations of MD technique see e.g. Refs. [49,55].

We have also searched for alternative structures consistent with the primary NMR data. We did not find any such structures in our simulations of the two-triad triplex. However, we also monitored violations of the NMR data for the upper two triads in the above-reported simulations of structures **E, F** and **J** (nucleotides G2, G3, G9, G10, G14, G15; in the molecule **J** the numbering is shifted by 7; Supporting Fig. S6 and Table SII). In this case we found two highly populated structures that would be consistent with the NMR data applicable for the G-triads. The first one is the ideal Hoogsteen triad stack, that could be either ion-stabilized or not. The second one is the common non-planar arrangement consisting of the Hoogsteen-paired duplex with the remaining strand tilted by ~80° and binding a cation in the perturbed channel (Supporting Fig. S7). We think that the existence of the latter arrangement could be supported by absence of any NMR correlation between G10 and any other triplex stem deoxyguanosine in the original experiment [50]. Therefore, it cannot be ruled out that the NMR data may also capture imperfect triplexes, although we were not able to stabilize such structures for the 15-TBA two-triad system in simulations (see the Supporting material – Section 10.5 – for further details). The primary NMR data in Ref. [50] might be insufficient to differentiate between ideally structured ion-stabilized triplex and some alternative structures with perturbed triads, and the suggested structures can to certain extent reflect the MD simulation data.

3.7. Triplex with fourth unbound G-strand did not show any attempts of folding into a quadruplex

We have simulated (5 μs) the structure **J** in an attempt to see if the unbound strand **a** would show some tendency to bind with the triplex. The simulation indicates that such rearrangements are unreachable on our affordable μs simulation time scale and no noticeable movement of the strand **a** is evidenced (Supporting material – Section 10.6). This result may even suggest that average time needed to capture the fourth strand may be longer than typical lifetime of the triplexes (at least as suggested by the

simulations). This, however, would not present any problem in participation of the triplexes in the folding pathway provided they are formed with sufficient probability.

3.8. Cation binding to narrow groove leads to increased structural stability

In all our simulations we have noticed that triplexes forming the narrow groove between two of their strands (molecules **C**, strands [**c** + **d**], **E** [**c** + **d**], **F** [**b** + **c**], **H** [**b** + **c**] and **J** [**c** + **d**]) incorporate cations from the bulk solvent into the narrow groove (cf. ref. [30] for explanation of the topological relation between strand orientation and type of G-DNA grooves). This can be enabled by the closer space between the phosphates creating deeper electrostatic potential minima inside the narrow groove and optimal spatial arrangement for ion binding. Once bound, the cation usually bridges the phosphates and stiffens the structure. The bridging can be direct between two phosphates (phosphate − sodium cation − phosphate) or water-mediated (phosphate − sodium cation − water molecule − phosphate), including up to three phosphates (Fig. 7). The direct binding is more stable (in terms of structural dynamics and residency time) and the cation resides there for dozens of nanoseconds before being replaced by a water and then by another cation. Moreover, long-residency structural water molecules are included. The water-mediated binding seems weaker (more dynamical), the cation is usually bound for several nanoseconds and then it is replaced by another one with a short period of several nanoseconds with only water molecules at the binding site. The narrow groove is able to accommodate more than one cation at the same time, however, this is rare. The direct and water-mediated ion binding patterns are interchangeable. The narrow groove adjusts its shape based on the type of cation binding pattern and number of phosphates included, i.e., it is narrower during the direct binding than during the water-mediated binding. When no cation is bound, the narrow groove is temporarily widened. Due to the dynamic fluctuations of the groove width, several distinguishable ion binding sites may appear along the groove, depending on the particular groove geometry.

The simulations suggest that the narrow groove cation binding leads to increased stability of the respective *c*WH duplexes. This could explain why the structure **C** is more structurally stable in simulations than **D**, despite the fact that both possess one lateral and one diagonal loop, and why structure **G** with one propeller and one lateral loop changes its structure more often than structure **H**. Strand relocations and shifts from the lateral to the diagonal loop of the structure **E** could be related to the larger stability of the narrow groove, since the strand relocations always disrupt duplexes topologically forming the wide groove. Although some perturbations (usually reversible) and strand relocations (one simulation of the molecule **F**) are seen also for the narrow groove, these structural changes always occur right after the start of the simulation, when the narrow groove does not yet develop its ion-binding pocket. Of course, we do not suggest that the ion binding fully freezes the narrow grooves. However, their dynamics is definitely attenuated and conformational changes are likely initiated in periods of ion exchange, when the narrow grooves temporarily lack the bound ion.

We suggest that the specific ion binding to the narrow grooves may be another structural and free energy factor which contributes to the basic rules determining stabilities of different G-DNA topologies [22,29,30,63]. We plan to further investigate this issue in the near future. We inspected all our available simulations of complete quadruplexes and found that cation binding site in the narrow grooves is a common feature, relevant for Na^+ as well as K^+ ions.

4. Discussion

We have carried out unbiased explicit solvent MD simulations of a set of G-triplexes which can participate in folding pathways of the human telomeric quadruplexes. Starting structures were obtained from the relevant experimental quadruplex structures by deleting either the first or the last strand (Fig. 1). Due to the nature of the starting structures, the present study characterizes properties of G-triplexes with native-like arrangements assuming that they are already folded. We have considered all possible triplex architectures that could be relevant to G-DNA folding as direct (i.e., with no need for further substantial rearrangements) intermediates. The 1–5 μs simulations are by three orders of magnitude longer compared to earlier such studies [43,44] and are done with the substantially improved parmbsc0 force field [76].

The μs time scale is essential since ns scale simulations remain very close to the starting configurations and do not show almost any dynamics. The unbiased simulations can investigate properties of diverse structures that can participate in the different stages and substates of the folding process. Obviously, the simulations cannot determine relative stabilities of these conformations and their relative populations during the folding, which in addition may substantially vary for different sequences and experimental conditions (including potential prefolding of the single strands). Although our simulations remain far from being converged, they nevertheless provide plausible estimate of lifetimes of the G-triplexes (as predicted by contemporary MD, see below), because fraction of our simulations achieves spontaneous loss of the triplexes. Obviously, we cannot rule out that the observed unfolding events might be intermediate steps necessary for the system to pass to another differently folded and more stable state on a time scale not affordable with contemporary computers. However, it does not have any effect on our conclusions, since all our triplex simulations started from the native geometries that would be needed for the G-DNA folding. So, even if the structures eventually fold in to something else, it would mean misfolded structures, which are not suitable for the folding of the G-DNA. Based on the theory of rare events [88] our simulations are qualitatively

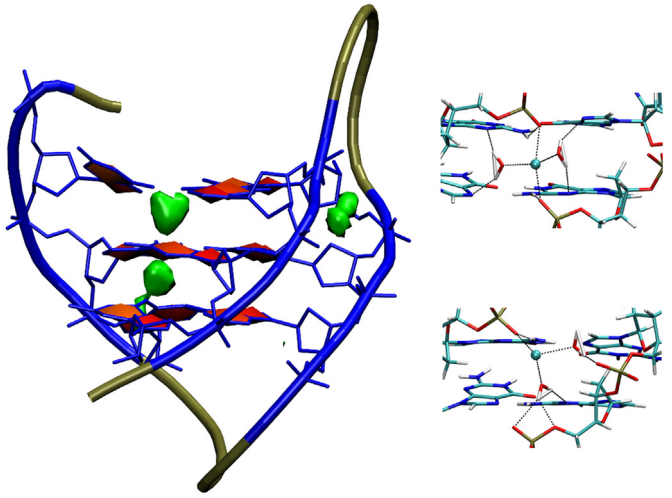

Fig. 7. Left: Increased sodium cation density (green) inside the triplex channel and the narrow groove. Deoxyguanosine residues are in blue, backbone is tan, hydrogen atoms are not shown. Top right: Sodium cation bridging two phosphates from adjacent G-strands directly across the narrow groove. Two structural water molecules are present. Bottom right: Water-mediated sodium-cation bridge between three phosphates from adjacent G-strands across the narrow groove. The cation is shown as cyan sphere, bonding is depicted by dashed black line.

consistent with triplex lifetimes in the μs range, and this result is discussed below.

Direct comparison of the simulations with the available experimental data is not straightforward, since the experiments do not provide details of the atomic structures occurring during the folding of the human telomeric quadruplexes. Many recent papers suggested presence of triplex intermediates, assuming either on-pathway or off-pathway roles of the triplexes [37,40–44,46], with various time scales. The experimental studies usually assume or resolve a limited number of intermediates [32–34,37]. The most comprehensive kinetics experiments were recently reported by Gray et al. [38] and suggested a very complex folding with the individual contributing processes having time scales from ms to 1000 s. Still, even such sophisticated experiments do not allow obtaining direct structural insights. Thus, the experiments and MD simulations should be considered as complementary approaches. The simulations available so far indirectly suggest a possibility that there can be a number of stable misfolded substates along the folding and unfolding pathways [48]. These states or structures would result in complex folding landscapes (perhaps a folding funnel). In fact, complex atomistic-level folding pathway has been demonstrated by unbiased simulations even for the simplest DNA system, a short hairpin loop [58]. The hairpin loop is the only nucleic acid molecule that has so far been spontaneously folded in unbiased simulations mimicking real dynamics of the molecules.

Essentially all folding models that have been so far suggested for the human telomeric quadruplex based on the experiments assume (at least as visualized by the figures) that the molecules adopt at the very beginning of the process the right combination of the *syn*/*anti* guanine orientations and then proceed further through straightforward intermediates towards the final fold. In contrast, the simulations indirectly suggest that, at the level of single molecules, the molecules can adopt numerous diverse structures with diverse *syn*/*anti* guanine combinations. No physicochemical mechanism which would eliminate such structures from the folding pathways has so far been suggested. While we cannot rule out that folding of some G-DNA sequences avoids substantial populations of stable structures with non-native *syn*/*anti* guanine distributions, we also cannot exclude that the experiments so far could not resolve their participation. Whether the present models are oversimplified due to the limited resolution of the experiments or whether the human telomeric DNA indeed eliminates the structures with nonnative *syn*/*anti* patterns so that they do not affect the folding process remains to be answered by further research.

4.1. Different G-triplexes have different life-times

The present simulations suggest that life-times of antiparallel triplexes (at least μs-scale, see below) with lateral and diagonal loops are sufficient to contribute to the folding pathways. The simulations, however, also show that the triplex states may involve structures with various local deviations from the ideal triplex pairing, such as strand tilting and deformed triads. The triplexes may, besides the ideal cation-stabilized Hoogsteen triads, adopt also alternative triads, including symmetrical triangle ones (Figs. 2 and 5). The simulations reveal easy rearrangements of the strands allowing interchanges between lateral and diagonal loops (Fig. 5, Supporting Figs. S3 and S8). Although our simulations did not directly show any transition from diagonal to lateral loops, we suggest that such movements are also easily accessible. Lack of such transitions in our simulations may merely reflect the sampling limitations. The fine balance in lateral loop ↔ diagonal loop transitions may also be sensitive to approximations of the simulation methodology. Transition from lateral to diagonal loop is easier when it includes loss of a wide groove, since narrow grooves seem to be stiffened by ion binding sites (Fig. 7).

Antiparallel triplexes with propeller loops appear to be somewhat less stable. The propeller loop may specifically interact with the G-strands and interfere with the triplex structure (Fig. 6). The propeller loops may also more easily unfold. Therefore, formation of propeller loops may complicate the folding pathways of the respective quadruplexes compared to the diagonal and lateral loops. Still, considering all the data, antiparallel G-triplexes with propeller loop are viable candidates to participate in the folding pathways.

Parallel triplexes with all guanines having *anti* conformation are quite unstable and therefore unlikely to contribute to the folding of the human telomeric quadruplex (Fig. 3). Their life-time appears to be insufficient to play any significant role in the folding. The parallel all-*anti* quadruplexes can form by the nucleation and strand-slippage mechanism described earlier [47,48]. Parallel triplexes were tentatively suggested to potentially occur in unfolding of parallel quadruplexes based on singular value decomposition analysis of CD melting data under dehydrating conditions, however, they were not suggested to participate in folding of the other human telomeric G-DNA topologies [37]. For the sake of completeness we note that we did not investigate potential parallel triplexes with mixed *syn* and *anti* guanine orientations. These triplexes were not studied since the currently known folds of the human telomeric quadruplexes do not include them. Such triplexes would be relevant for example in folding of a human telomeric structure having two propeller loops followed by a lateral loop [30], which most likely would have 5′-*syn*-*anti*-*anti*-3′ pattern in all three strands [63].

4.2. The sampling limitation

In order to properly understand the present results, it is important to consider the limitations of the method. The first limitation is the time-scale of the simulations. As reviewed elsewhere [49], contemporary simulations are equivalent to hypothetical microsecond-scale single molecule experiments where the molecule initially adopts some starting configuration and subsequently undergoes spontaneous unbiased structural dynamics. In other words, our simulations assess the structural stability of potential G-triplex structures and their lifetimes starting from the folded structure. The simulations do not capture stability in a thermodynamics free energy sense.

Our simulations quite confidently prove that antiparallel G-triplexes would have at least μs life-times sufficient to participate in the folding. However, our results are not sufficient to prove that the G-triplexes form during the quadruplex folding pathways with sufficient populations. We also cannot rule out the possibility that the G-triplexes are out-competed by other structures (not included in our study) during the folding process. All simulations started from folded triplexes with the correct (native) distribution of the *syn*/*anti* orientations of the guanines. We did not assess potential other structures, mainly misfolded four stranded structures with non-native distribution of the *syn*/*anti* orientations of the guanines (see for example ref. [48]). Recent simulations revealed that such misfolded quadruplexes, once formed, would have very long life-times (likely longer than the G-triplexes) and we do not see any obvious reason why they should not be populated during the process of G-DNA folding at least as off-pathway structures [48]. Contemporary computation methods do not provide any reliable tool to evaluate relative free energies of G-triplexes and misfolded quadruplexes. Nevertheless, if we compare molecular interactions in G-triplexes and misfolded quadruplexes with shifted strands [48], the misfolded quadruplexes should have, once formed, longer lifetimes (smaller $k_{unfolding}$) than the triplexes. The G-triplex

contains three stacked triads while the misfolded quadruplex could be locked by two full ion-stabilized tetrads further supported by one stacked triad (Fig. 8). Actually, the 2KF8 structure adopts a shifted-strand topology.

4.3. Force field limitation and the life-time of the triplexes: microseconds, milliseconds or hundreds of seconds?

Assessment of the effect of the force field approximation is complex. The parmbsc0 AMBER force field provides very good description of quadruplex stems [48,78] with less accurate description of the loop regions [78,89]. Loops are, however, not a major concern for our study, as the main conclusions are based on qualitative analysis of structural dynamics and life-times of different folds of G-triplexes. We have also tested the latest parmbsc0$\chi_{OL4}\varepsilon\zeta_{OL1}$ refinement of the force field [78,79]. It is encouraging to see that this DNA force field variant recommended by the latest version of the AMBER program package predicts either the same or even further increased stability of the studied G-triplexes compared to parmbsc0. Thus, all recent versions of the force field provide consistent results. Definitely, regarding the structural aspects described in our study, sampling (length of the simulations) is more limiting than the choice of the force field variant.

Second issue is related to the comparison of the lifetime of the G-triplexes observed by simulations and suggested by the various experimental models. There are indications that nucleic acids simulations in some cases underestimate the actual stability of base-paired regions of nucleic acids structures [90,91]. Thus we assume (although we do not have any direct proof of that) that the absolute structural stability (life-times) of G-triplexes in our simulations might be underestimated. The reason why force fields may underestimate the stability of G-triplexes is the lack of explicit inclusion of polarization and electronic-structure redistribution effects which contribute to the stabilization of H-bonded base pairs and to the ion binding. For example, RNA stem-tetraloop systems are stable in solution experiments with just two base pairs in the canonical A-RNA stems [92]. In contrast, solution simulations rather include three base pairs in the stem to prevent excessive end fraying [56,93]. Similarly, in simulations of DNA and RNA oligonucleotide duplexes, the terminal base pairs are often unstable due to the end fraying [90,94]. Although the end fraying is a real effect present in equivalent solution experiments, the simulations probably overestimate its magnitude [90]. Further, the simulations may underestimate binding of the ions to the G-triplex channel due to the pair-additive nature of the force field [95,96]. In summary, we hypothesize that real G-triplexes could have longer life-times than indicated by our simulations. Any such increase in stability would further support plausibility of the G-triplex intermediates. This effect would be identical for all force field versions used in this study since it stems from the basic approximation of the non-polarizable force field. Relative stability ranking (relative life-times) of the different types of three-triad triplexes should be correctly captured by the simulations. Life-time and stability of the two-tetrad triplex could be affected to a larger extent. Future well-tuned polarizable force fields may provide additional insights into the stability of the G-triplexes [97].

A modest force field underestimation of the triplex stability could lead to underestimation of their lifetimes by say 2-3 orders of magnitude. This would shift the G-triplex lifetimes to the millisecond range compared to the rather microsecond range suggested by the present simulation data. Nevertheless, this would still not be sufficient to make the computations consistent with some experimental models, since for example the most advanced kinetic studies suggest models with triplexes having lifetimes of hundreds of seconds [38]. Such lifetimes of triplexes are already difficult to reconcile with the simulation data without assuming a serious inaccuracy of the force field description. Nevertheless, even this possibility cannot be ruled out. Thus, further research is vital to clarify if either the current experimental models of very stable triplexes are somehow over-simplified or if these models are correct and the underestimation of the lifetime of the G-triplexes by contemporary force fields is so large.

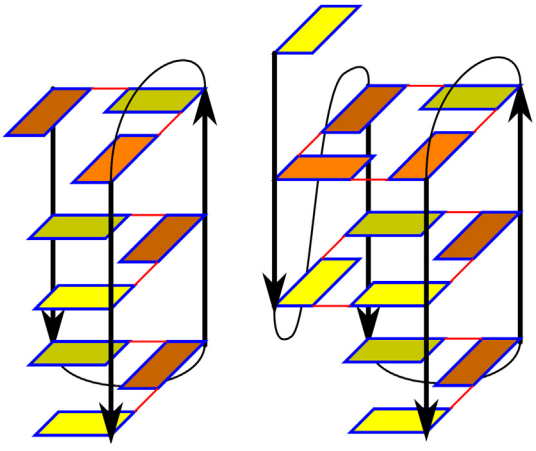

Fig. 8. Comparison of a triplex (left) with a misfolded quadruplex with strand slippage (right). The triplex contains three stacked triads, while the misfolded quadruplex is made of two consecutive stacked tetrads with a triad, which should lead to longer life-time compared to the triplex. Note that ions are not shown. Structural schemes are visualized as in Fig. 1.

4.4. The 15-TBA two-triad structure simulations

We have also carried out simulations of a two-triad triplex, which has been suggested as a folding intermediate of the 15-TBA quadruplex with a direct support by NMR [50]. Our simulations suggest that such triplexes should have lifetimes sufficient to contribute to the folding process. The proposed structure remained intact in ~30–40% of our 1 μs simulations (Supporting Table SI). Nevertheless, in contrast to the original study [50], sufficiently long simulations would probably not predict the triplex to be the dominantly populated state of the truncated 15-TBA sequence used in the NMR experiment. However, the two-triad triplex can be particularly sensitive to the above-noted “end-effects” and thus the force field can underestimate its free energy more than in case of the three-triad structures. As discussed in Section 3.6, it can reconcile the difference between the MD and NMR data. The primary experimental NMR data are consistent with structure of the ideal Hoogsteen triplex with a coordinated ion [50]. We further show that the NMR restraints (as far as definition of the triads is concerned) could be, in principle, also consistent with deformed triplex with one strand tilted (Fig. S7). Unfortunately, with presently available time scales and force fields we are unable to fully explore the free energy surface of the molecule so that further experimental data would be vital. It should also be noted that results obtained for the 15-TBA system are not necessarily fully transferable to the human telomeric quadruplex folding. Folding routes of two-tetrad and three-tetrad quadruplexes can be quite different, since folding pathways of two-tetrad systems should not be affected by off-pathway two-tetrad intermediates with slipped strands that can be easily stabilized for three-tetrad quadruplexes [48]. As discussed in the

Introduction (Section 1), the pool of potential intermediates expands dramatically with increasing the number of guanines in the G-stretches and the variable combinations of *syn*/*anti* orientations of the guanines in the individual strands.

5. Conclusions

We demonstrate that although contemporary simulations remain far from providing a complete description of the folding pathway of the human telomeric quadruplexes, they provide insights into selected aspects of the folding processes. We show beyond any doubt that G-triplexes are quite stable *per se* and can play a role during the folding. The atomistic picture of the folding via triplex pathway could be quite complex and may include competition between Hoogsteen and alternative triads, between lateral and diagonal loops, perturbed structures with strand inclination, etc. However, we cannot rule out that the folding also involves other structures that are even more stable than the G-triplexes, such as quadruplexes with shifted strands.

Conflict of interest

None declared.

Acknowledgment

This work was supported by the Czech Science Foundation [grant numbers P208/11/1822, 13-28310S]; the project "CEITEC - Central European Institute of Technology" [CZ.1.05/1.1.00/02.0068] from European Regional Development Fund; LT was supported by Marie Curie re-integration program (ECOPOD) and European Molecular Biology Organization [IG 2535]; T.E.C. acknowledges funding from National Institutes of Health [R01-GM081411]; access to storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" [LM2010005], is greatly appreciated. The computations were done exclusively using equipment funded by the CEITEC project.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.biochi.2014.07.009.

References

[1] A.T. Phan, Human telomeric G-quadruplex: structures of DNA and RNA sequences, Febs J. 277 (2010) 1107–1117.
[2] E.Y.N. Lam, D. Beraldi, D. Tannahill, S. Balasubramanian, G-quadruplex structures are stable and detectable in human genomic DNA, Nat. Commun. 4 (2013) 1796.
[3] G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, Quantitative visualization of DNA G-quadruplex structures in human cells, Nat. Chem. 5 (2013) 182–186.
[4] R. Haensel, S. Foldynova-Trantirkova, F. Loehr, J. Buck, E. Bongartz, E. Bamberg, H. Schwalbe, V. Doetsch, L. Trantirek, Evaluation of parameters critical for observing nucleic acids inside living Xenopus laevis oocytes by in-cell NMR spectroscopy, J. Am. Chem. Soc. 131 (2009) 15761–15768.
[5] H.J. Lipps, D. Rhodes, G-quadruplex structures: in vivo evidence and function, Trends Cell. Biol. 19 (2009) 414–422.
[6] A. Rizzo, E. Salvati, M. Porru, C. D'Angelo, M.F. Stevens, M. D'Incalci, C. Leonetti, E. Gilson, G. Zupi, A. Biroccio, Stabilization of quadruplex DNA perturbs telomere replication leading to the activation of an ATR-dependent ATM signaling pathway, Nucleic Acids Res. 37 (2009) 5353–5364.
[7] A. Mendez-Bermudez, M. Hills, H.A. Pickett, A.T. Phan, J.-L. Mergny, J.-F. Riou, N.J. Royle, Human telomeres that contain (CTAGGG)(n) repeats show replication dependent instability in somatic cells and the male germline, Nucleic Acids Res. 37 (2009) 6225–6238.
[8] S.A. Juranek, K. Paeschke, Cell cycle regulation of G-quadruplex DNA structures at telomeres, Curr. Pharm. Des. 18 (2012) 1867–1872.
[9] J. Postberg, M. Tsytlonok, D. Sparvoli, D. Rhodes, H.J. Lipps, A telomerase-associated RecQ protein-like helicase resolves telomeric G-quadruplex structures during replication, Gene 497 (2012) 147–154.
[10] S. Neidle, Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer, Febs J. 277 (2010) 1118–1125.
[11] M. Duechler, G-quadruplexes: targets and tools in anticancer drug design, J. Drug. Target. 20 (2012) 389–400.
[12] K.W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D.J. Patel, K.N. Luu, A.T. Phan, Structure of the human telomere in K+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers, J. Am. Chem. Soc. 131 (2009) 4301–4309.
[13] Y. Wang, D.J. Patel, Solution structure of the human telomeric repeat d [AG(3)(T(2)AG(3))3] G-tetraplex, Structure 1 (1993) 263–282.
[14] G.N. Parkinson, M.P.H. Lee, S. Neidle, Crystal structure of parallel quadruplexes from human telomeric DNA, Nature 417 (2002) 876–880.
[15] A. Ambrus, D. Chen, J.X. Dai, T. Bialis, R.A. Jones, D.Z. Yang, Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution, Nucleic Acids Res. 34 (2006) 2723–2735.
[16] J. Dai, M. Carver, C. Punchihewa, R.A. Jones, D. Yang, Structure of the hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence, Nucleic Acids Res. 35 (2007) 4927–4940.
[17] R. Haensel, F. Loehr, S. Foldynova-Trantirkova, E. Bamberg, L. Trantirek, V. Doetsch, The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions, Nucleic Acids Res. 39 (2011) 5768–5775.
[18] R. Haensel, F. Loehr, L. Trantirek, V. Doetsch, High-resolution insight into G-overhang architecture, J. Am. Chem. Soc. 135 (2013) 2816–2824.
[19] S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, S. Neidle, Quadruplex DNA: sequence, topology and structure, Nucleic Acids Res. 34 (2006) 5402–5415.
[20] S. Neidle, The structures of quadruplex nucleic acids and their drug complexes, Curr. Opin. Struct. Biol. 19 (2009) 239–250.
[21] B. Heddi, A.T. Phan, Structure of human telomeric DNA in crowded solution, J. Am. Chem. Soc. 133 (2011) 9824–9833.
[22] M.W. da Silva, Geometric formalism for DNA quadruplex folding, Chem. Eur. J. 13 (2007) 9738–9745.
[23] M. Crnugelj, P. Sket, J. Plavec, Small change in a G-rich sequence, a dramatic change in topology: new dimeric G-quadruplex folding motif with unique loop orientations, J. Am. Chem. Soc. 125 (2003) 7866–7871.
[24] A.T. Phan, Y.S. Modi, D.J. Patel, Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter, J. Am. Chem. Soc. 126 (2004) 8710–8716.
[25] K.N. Luu, A.T. Phan, V. Kuryavyi, L. Lacroix, D.J. Patel, Structure of the human telomere in K+ solution: an intramolecular (3+1) G-quadruplex scaffold, J. Am. Chem. Soc. 128 (2006) 9963–9970.
[26] J.X. Dai, M. Carver, D.Z. Yang, Polymorphism of human telomeric quadruplex structures, Biochimie 90 (2008) 1172–1183.
[27] K.W. Lim, V.C.M. Ng, N. Martín-Pintado, B. Heddi, A.T. Phan, Structure of the human telomere in Na+ solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity, Nucleic Acids Res. 41 (2013) 10556–10562.
[28] P. Agrawal, E. Hatzakis, K. Guo, M. Carver, D. Yang, Solution structure of the major G-quadruplex formed in the human VEGF promoter in K+: insights into loop interactions of the parallel G-quadruplexes, Nucleic Acids Res. 41 (2013) 10584–10592.
[29] X.H. Cang, J. Sponer, T.E. Cheatham, Explaining the varied glycosidic conformational, G-tract length and sequence preferences for anti-parallel G-quadruplexes, Nucleic Acids Res. 39 (2011) 4499–4512.
[30] A.I. Karsiotis, C. O'Kane, M.W. da Silva, DNA quadruplex folding formalism – a tutorial on quadruplex topologies, Methods 64 (2013) 28–35.
[31] Y. Xue, J.-q. Liu, K.-w. Zheng, Z.-y. Kan, Y.-h. Hao, Z. Tan, Kinetic and thermodynamic control of G-quadruplex folding, Angew. Chem. Int. Ed. 50 (2011) 8046–8050.
[32] A.Y.Q. Zhang, S. Balasubramanian, The kinetics and folding pathways of intramolecular G-quadruplex nucleic acids, J. Am. Chem. Soc. 134 (2012) 19297–19308.
[33] R.D. Gray, J.B. Chaires, Kinetics and mechanism of K(+)- and Na(+)-induced folding of models of human telomeric DNA into G-quadruplex structures, Nucleic Acids Res. 36 (2008) 4191–4203.
[34] R.D. Gray, L. Petraccone, J.O. Trent, J.B. Chaires, Characterization of a K(+)-induced conformational switch in a human telomeric DNA oligonucleotide using 2-aminopurine fluorescence, Biochemistry 49 (2010) 179–194.
[35] A. Marchand, R. Ferreira, H. Tateishi-Karimata, D. Miyoshi, N. Sugimoto, V. Gabelica, Sequence and solvent effects on telomeric DNA bimolecular G-quadruplex folding kinetics, J. Phys. Chem. B 117 (2013) 12391–12401.
[36] I. Fotticchia, C. Giancola, L. Petraccone, G-quadruplex unfolding in higher-order DNA structures, Chem. Commun. 49 (2013) 9488–9490.
[37] R. Buscaglia, R.D. Gray, J.B. Chaires, Thermodynamic characterization of human telomere quadruplex unfolding, Biopolymers 99 (2013) 1006–1018.
[38] R.D. Gray, J.O. Trent, J.B. Chaires, Folding and unfolding pathways of the human telomeric G-quadruplex, J. Mol. Biol. 426 (2014) 1629–1650.

[39] Y. Li, C. Liu, X.J. Feng, Y.Z. Xu, B.F. Liu, Ultrafast microfluidic mixer for tracking the early folding kinetics of human telomere G-quadruplex, Anal. Chem. 86 (2014) 4333–4339.
[40] M. Boncina, J. Lah, I. Prislan, G. Vesnaver, Energetic basis of human telomeric DNA folding into G-quadruplex structures, J. Am. Chem. Soc. 134 (2012) 9657–9663.
[41] W. Li, X.-M. Hou, P.-Y. Wang, X.-G. Xi, M. Li, Direct measurement of sequential folding pathway and energy landscape of human telomeric G-quadruplex structures, J. Am. Chem. Soc. 135 (2013) 6423–6426.
[42] R.D. Gray, R. Buscaglia, J.B. Chaires, Populated intermediates in the thermal unfolding of the human telomeric quadruplex, J. Am. Chem. Soc. 134 (2012) 16834–16844.
[43] T. Mashimo, H. Yagi, Y. Sannohe, A. Rajendran, H. Sugiyama, Folding pathways of human telomeric type-1 and type-2 G-quadruplex structures, J. Am. Chem. Soc. 132 (2010) 14910–14918.
[44] D. Koirala, T. Mashimo, Y. Sannohe, Z.B. Yu, H.B. Mao, H. Sugiyama, Intramolecular folding in three tandem guanine repeats of human telomeric DNA, Chem. Commun. 48 (2012) 2006–2008.
[45] A. Rajendran, M. Endo, K. Hidaka, H. Sugiyama, Direct and single-molecule visualization of the solution-state structures of G-hairpin and G-triplex intermediates, Angew. Chem. 126 (2014) 4191–4196.
[46] D. Koirala, C. Ghimire, C. Bohrer, Y. Sannohe, H. Sugiyama, H.B. Mao, Long-loop G-quadruplexes are misfolded population minorities with fast transition kinetics in human telomeric sequences, J. Am. Chem. Soc. 135 (2013) 2235–2241.
[47] R. Stefl, T.E. Cheatham, N. Spackova, E. Fadrna, I. Berger, J. Koca, J. Sponer, Formation pathways of a guanine-quadruplex DNA revealed by molecular dynamics and thermodynamic analysis of the substates, Biophys. J. 85 (2003) 1787–1804.
[48] P. Stadlbauer, M. Krepl, T.E. Cheatham, J. Koca, J. Sponer, Structural dynamics of possible late-stage intermediates in folding of quadruplex DNA studied by molecular simulations, Nucleic Acids Res. 41 (2013) 7128–7143.
[49] J. Sponer, X.H. Cang, T.E. Cheatham, Molecular dynamics simulations of G-DNA and perspectives on the simulation of nucleic acid structures, Methods 57 (2012) 25–39.
[50] V. Limongelli, S. De Tito, L. Cerofolini, M. Fragai, B. Pagano, R. Trotta, S. Cosconati, L. Marinelli, E. Novellino, I. Bertini, A. Randazzo, C. Luchinat, M. Parrinello, The G-triplex DNA, Angew. Chem. Int. Ed. 52 (2013) 2269–2273.
[51] H. Li, E.H. Cao, T. Gisler, Force-induced unfolding of human telomeric G-quadruplex: a steered molecular dynamics simulation study, Biochem. Biophys. Res. Commun. 379 (2009) 70–75.
[52] C. Yang, S. Jang, Y. Pak, Multiple stepwise pattern for potential of mean force in unfolding the thrombin binding aptamer in complex with Sr2+, J. Chem. Phys. 135 (2011) 225104.
[53] E. Kim, C. Yang, Y. Pak, Free-energy landscape of a thrombin-binding DNA aptamer in aqueous environment, J. Chem. Theory Comput. 8 (2012) 4845–4851.
[54] Y. Bian, C. Tan, J. Wang, Y. Sheng, J. Zhang, W. Wang, Atomistic picture for the folding pathway of a hybrid-1 type human telomeric DNA G-quadruplex, PLoS Comput. Biol. 10 (2014) e1003562.
[55] J. Sponer, P. Banas, P. Jurecka, M. Zgarbova, P. Kuehrova, M. Havrila, M. Krepl, P. Stadlbauer, M. Otyepka, Molecular dynamics simulations of nucleic acids. From tetranucleotides to the ribosome, J. Phys. Chem. Lett. 5 (2014) 1771–1782.
[56] P. Kuehrova, P. Banas, R.B. Best, J. Sponer, M. Otyepka, Computer folding of RNA tetraloops? Are we there yet? J. Chem. Theory Comput. 9 (2013) 2115–2125.
[57] R.V. Reshetnikov, J. Sponer, O.I. Rassokhina, A.M. Kopylov, P.O. Tsvetkov, A.A. Makarov, A.V. Golovin, Cation binding to 15-TBA quadruplex DNA is a multiple-pathway cation-dependent process, Nucleic Acids Res. 39 (2011) 9789–9802.
[58] G. Portella, M. Orozco, Multiple routes to characterize the folding of a small DNA hairpin, Angew. Chem. Int. Ed. 49 (2010) 7673–7676.
[59] A. Sali, E. Shakhnovich, M. Karplus, How does a protein fold? Nature 369 (1994) 248–251.
[60] J.D. Bryngelson, J.N. Onuchic, N.D. Socci, P.G. Wolynes, Funnels, pathways, and the energy landscape of protein-folding– a synthesis, Proteins: Struct. Funct. Genet. 21 (1995) 167–195.
[61] K.A. Dill, S.B. Ozkan, M.S. Shell, T.R. Weikl, The protein folding problem, Annu. Rev. Biophys. Annu. Rev. Palo Alto (2008) 289–316.
[62] D. Thirumalai, C. Hyeon, RNA and protein folding: common themes and variations, Biochemistry 44 (2005) 4957–4970.
[63] J. Sponer, A. Mladek, N. Spackova, X.H. Cang, T.E. Cheatham, S. Grimme, Relative stability of different DNA guanine quadruplex stem topologies derived using large-scale quantum-chemical computations, J. Am. Chem. Soc. 135 (2013) 9785–9796.
[64] B. Islam, M. Sgobba, C. Laughton, M. Orozco, J. Sponer, S. Neidle, S. Haider, Conformational dynamics of the human propeller telomeric DNA quadruplex on a microsecond time scale, Nucleic Acids Res. 41 (2013) 2723–2735.
[65] R. Reshetnikov, A. Golovin, V. Spiridonova, A. Kopylov, J. Sponer, Structural dynamics of thrombin-binding DNA aptamer d(GGTTGGTGTGGTTGG) quadruplex DNA studied by large-scale explicit solvent simulations, J. Chem. Theory Comput. 6 (2010) 3003–3014.
[66] H. Zhu, S. Xiao, H. Liang, Structural dynamics of human telomeric G-quadruplex loops studied by molecular dynamics simulations, PLoS One 8 (2013) e71380.
[67] P. Schultze, R.F. Macaya, J. Feigon, 3-dimensional solution structure of the thrombin-binding DNA Aptamer d(GGTTGGTGTGGTTGG), J. Mol. Biol. 235 (1994) 1532–1547.
[68] N. Guex, M.C. Peitsch, Swiss-model and the Swiss-PdbViewer: an environment for comparative protein modeling, Electrophoresis 18 (1997) 2714–2723.
[69] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.
[70] W.L. Jorgensen, Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers – application to liquid water, J. Am. Chem. Soc. 103 (1981) 335–340.
[71] H.J.C. Berendsen, J.R. Grigera, T.P. Straatsma, The missing term in effective pair potentials, J. Phys. Chem. 91 (1987) 6269–6271.
[72] I.S. Joung, T.E. Cheatham, Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations, J. Phys. Chem. B 112 (2008) 9020–9041.
[73] J.B. Chaires, Human telomeric G-quadruplex: thermodynamic and kinetic studies of telomeric quadruplex stability, Febs J. 277 (2010) 1098–1106.
[74] R.D. Gray, J. Li, J.B. Chaires, Energetics and kinetics of a conformational switch in G-quadruplex DNA, J. Phys. Chem. B 113 (2009) 2676–2683.
[75] D.A.D. Case T, T.E. Cheatham III, C.L. Simmerling, J.D. Wang R.E., R. Luo, R. Walker, W. Zhang K.M.R, S. Merz, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, K.F. I.W. Kolossvai, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu S.R.S, T. Brozell, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, V.L.C. Sagui, T. Babin, S. Gusarov, A. Kovalenko, P.A. Kollman, AMBER 12, University of California, San Francisco, 2012.
[76] A. Perez, I. Marchan, D. Svozil, J. Sponer, T.E. Cheatham, C.A. Laughton, M. Orozco, Refinenement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers, Biophys. J. 92 (2007) 3817–3829.
[77] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197.
[78] M. Krepl, M. Zgarbova, P. Stadlbauer, M. Otyepka, P. Banas, J. Koca, T.E. Cheatham, P. Jurecka, J. Sponer, Reference simulations of noncanonical nucleic acids with different chi variants of the AMBER force field: quadruplex DNA, quadruplex RNA, and Z-DNA, J. Chem. Theory Comput. 8 (2012) 2506–2520.
[79] M. Zgarbova, F.J. Luque, J. Sponer, T.E. Cheatham, M. Otyepka, P. Jurecka, Toward improved description of DNA backbone: revisiting epsilon and zeta torsion force field parameters, J. Chem. Theory Comput. 9 (2013) 2339–2354.
[80] A.W. Goetz, M.J. Williamson, D. Xu, D. Poole, S. Le Grand, R.C. Walker, Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born, J. Chem. Theory Comput. 8 (2012) 1542–1555.
[81] S. Le Grand, A.W. Goetz, R.C. Walker, SPFP: speed without compromise – a mixed precision model for GPU accelerated molecular dynamics simulations, Comput. Phys. Commun. 184 (2013) 374–380.
[82] R. Salomon-Ferrer, A.W. Goetz, D. Poole, S. Le Grand, R.C. Walker, Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald, J. Chem. Theory Comput. 9 (2013) 3878–3888.
[83] T. Darden, D. York, L. Pedersen, Particle mesh Ewald - an N.log(N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.
[84] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577–8593.
[85] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of cartesian equations of motion of a system with constraints - molecular dynamics of n-alkans, J. Comput. Phys. 23 (1977) 327–341.
[86] H.J.C. Berendsen, J.P.M. Postma, W.F. Vangunsteren, A. Dinola, J.R. Haak, Molecular-dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.
[87] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics, J. Mol. Graph. Model. 14 (1996) 33–38.
[88] P. Hanggi, P. Talkner, M. Borkovec, Reaction-rate theory – 50 years after Kramers, Rev. Mod. Phys. 62 (1990) 251–341.
[89] E. Fadrna, N. Spackova, J. Sarzynska, J. Koca, M. Orozco, T.E. Cheatham, T. Kulinski, J. Sponer, Single stranded loops of quadruplex DNA as key benchmark for testing nucleic acids force fields, J. Chem. Theory Comput. 5 (2009) 2514–2530.
[90] T. Drsata, A. Perez, M. Orozco, A.V. Morozov, J. Sponer, F. Lankas, Structure, stiffness and substates of the Dickerson-Drew dodecamer, J. Chem. Theory Comput. 9 (2013) 707–721.
[91] S. Piana, Atomistic simulation of the DNA helix – coil transition, J. Phys. Chem. A 111 (2007) 12349–12354.
[92] S. Nozinovic, B. Furtig, H.R.A. Jonker, C. Richter, H. Schwalbe, High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA, Nucleic Acids Res. 38 (2010) 683–694.

[93] I. Besseova, P. Banas, P. Kuehrova, P. Kosinova, M. Otyepka, J. Sponer, Simulations of A-RNA duplexes. The effect of sequence, solute force field, water model, and salt concentration, J. Phys. Chem. B 116 (2012) 9899–9916.
[94] M. Zgarbova, M. Otyepka, J. Sponer, F. Lankas, P. Jurecka, Base pair fraying in molecular dynamics simulations of DNA and RNA, J. Chem. Theory Comput. (2014), http://dx.doi.org/10.1021/ct500120v.
[95] J. Song, C. Ji, J.Z.H. Zhang, The critical effect of polarization on the dynamical structure of guanine quadruplex DNA, Phys. Chem. Chem. Phys. 15 (2013) 3846–3854.
[96] K. Gkionis, H. Kruse, J.A. Platts, A. Mladek, J. Koca, J. Sponer, Ion binding to quadruplex DNA stems. Comparison of MM and QM descriptions reveals sizable polarization effects not included in contemporary simulations, J. Chem. Theory Comput. 10 (2014) 1326–1340.
[97] A. Savelyev, A.D. MacKerell Jr., All-atom polarizable force field for DNA based on the classical Drude oscillator model, J. Comput. Chem. 35 (2014) 1219–1239.